Spine

SPINE Volume 44, Number 16, pp 1144–1153

DEFORMITY

Development of Deployable Predictive Models for Minimal Clinically Important Difference Achievement Across the Commonly Used Health-related Quality of Life Instruments in Adult Spinal Deformity Surgery

Christopher P. Ames, MD,* Justin S. Smith, MD, PhD,† Ferran Pellisé, MD,‡ Michael P. Kelly, MD,§ Jeffrey L. Gum, MD,¶ Ahmet Alanay, MD,|| Emre Acaroğlu, MD,** Francisco Javier Sánchez Pérez-Grueso, MD,†† Frank S. Kleinstück, MD,‡‡ Ibrahim Obeid, MD,§§ Alba Vila-Casademunt, MSc,¶¶ Douglas C. Burton, MD,|||| Virginie Lafage, PhD,*** Frank J. Schwab, MD,*** Christopher I. Shaffrey, Sr, MD,† Shay Bess, MD,††† and Miquel Serra-Burriel, PhD‡‡‡, European Spine Study Group, International Spine Study Group

Study Design. Retrospective analysis of prospectively-collected, multicenter adult spinal deformity (ASD) databases.

From the *Department of Neurosurgery, University of California San Francisco, San Francisco, CA; †Department of Neurosurgery, University of Virginia Medical Center, Charlottesville, VA; ‡Spine Surgery Unit, Hospital Vall d'Hebron, Barcelona, Spain; §Department of Orthopaedic Surgery, Washington University, St Louis, MO; ¶Norton Leatherman Spine Center, Louisville, KY; ||Department of Orthopedics and Traumatology, Acibadem University, Büyükdere cad, Istanbul, Turkey; **Ankara ARTES Spine Center, Ankara, Turkey; ††Spine Surgery Unit, Hospital Universitario La Paz, Paseo de la Castellana, Madrid, Spain; ‡‡Spine Center Division, Department of Orthopedics and Neurosurgery, Schulthess Klinik, Zürich, Switzerland; §§Spine Surgery Unit, Bordeaux University Hospital, Bordeaux, France; ¶¶Vall d'Hebron Institute of Research (VHIR) Barcelona, Spain; ||||Department of Orthopaedic Surgery, University of Kansas Medical Center, Kansas City, KS; ***Department of Orthopaedic Surgery, Hospital for Special Surgery, New York, NY; †††Denver Denver International Spine Center, Presbyterian St. Luke's/Rocky Mountain Hospital for Children, Denver, CO; and ‡‡‡Center for Research in Health and Economics, Universitat Pompeu Fabra, Barcelona, Spain.

Acknowledgment date: December 3, 2018. Acceptance date: February 12, 2019.

The manuscript submitted does not contain information about medical device(s)/drug(s).

The International Spine Study Group Foundation receives funding support from DePuy Synthes, K2 M, Nuvasive, Orthofix, and Zimmer Biomet. The European Spine Study Group receives funding support from DePuy Synthes and Medtronic. Additional support was provided through Project PI16/01283, funded by Instituto de Salud Carlos III and co-funded by European Union (ERDF/ESF).

Relevant financial activities outside the submitted work: board membership, consultancy, royalties, grants, stocks, payment for lecture.

Address correspondence and reprint requests to Miquel Serra-Burriel, PhD, Center for Research in Health Economics, Department of Economics and Business, Universitat Pompeu Fabra, Office 23.111 Merce Rodoreda Building (Ciutadella Campus), Ramon Trias Fargas, 25-27, Barcelona 08005, Spain; E-mail: miquel.serrab@upf.edu

DOI: 10.1097/BRS.0000000000003031

Objective. To predict the likelihood of reaching minimum clinically important differences in patient-reported outcomes after ASD surgery.

Summary of Background Data. ASD surgeries are costly procedures that do not always provide the desired benefit. In some series only 50% of patients achieve minimum clinically important differences in patient-reported outcomes (PROs). Predictive modeling may be useful in shared-decision making and surgical planning processes. The goal of this study was to model the probability of achieving minimum clinically important differences change in PROs at 1 and 2 years after surgery.

Methods. Two prospective observational ASD cohorts were queried. Patients with Scoliosis Research Society-22, Oswestry Disability Index , and Short Form-36 data at preoperative baseline and at 1 and 2 years after surgery were included. Seventy-five variables were used in the training of the models including demographics, baseline PROs, and modifiable surgical parameters. Eight predictive algorithms were trained at four-time horizons: preoperative or postoperative baseline to 1 year and preoperative or postoperative baseline to 2 years. External validation was accomplished via an 80%/20% random split. Five-fold cross validation within the training sample was performed. Precision was measured as the mean average error (MAE) and R^2 values.

Results. Five hundred seventy patients were included in the analysis. Models with the lowest MAE were selected; R^2 values ranged from 20% to 45% and MAE ranged from 8% to 15% depending upon the predicted outcome. Patients with worse preoperative baseline PROs achieved the greatest mean improvements. Surgeon and site were not important components of the models, explaining little variance in the predicted 1- and 2-year PROs.

Conclusion. We present an accurate and consistent way of predicting the probability for achieving clinically relevant improvement after ASD surgery in the largest-to-date prospective operative multicenter cohort with 2-year follow-up. This study has significant clinical implications for shared decision making, surgical planning, and postoperative counseling.

Key words: adult spinal deformity surgery, MCID, predictive modeling, prognosis, shared decision-making.
Level of Evidence: 4
Spine 2019;44:1144–1153

In an era of increasing healthcare costs and progressive budget constraints, there is demand for providing value in care.[1,2] Adult spinal deformity (ASD) surgeries are costly and are associated with high complication rates and a frequent need for unanticipated reoperations, which further increase cost, making delivery of value-based care challenging. Analyses of cost-drivers in ASD surgery have found surgical techniques and instrumentation are among the potentially modifiable factors associated with increased cost of care. As value-driven healthcare evolves in ASD surgery, identification of other potentially modifiable factors is necessary to optimize care for the individual, including surgical techniques and approaches.

Patient-reported outcomes (PROs) tools are the common measure of outcome in value-driven health care. PRO instruments range from general measures of health, such as the Short-Form 36 (SF-36) to disease-specific measures, such as the Oswestry Disability Index (ODI) for lumbar disease and the Scoliosis Research Society-22r (SRS-22r) questionnaire for spinal deformity. The SRS-22r instrument is a composite of pain, self-image, function, and mental health domains; domains deemed most relevant to ASD patients. Historically, radiographic characteristics, specifically those related to sagittal plane alignment, have been the target for maximizing outcome improvement.[3] PROs are an subjective measure of health and patient expectations may be reflected in the reported scores and are not immune to well-documented bias such as anchoring.[4,5] Not surprisingly, self-image and mental health are associated with dissatisfaction and detrimental change in ASD, indicating a need for improved understanding of patient-level factors associated with outcomes improvement.[6,7] The minimal clinically important difference (MCID) is proposed as a threshold for surgical success, as it is the smallest change in a PRO measure deemed important to patients.[8]

Predictive modeling is a critical component of supply chain methodology in essentially all areas of consumerism. In a value-driven healthcare economy, delivery of medical care may be viewed from a similar perspective. Analytic techniques for modeling of large datasets in healthcare offer the opportunity to identify high-risk and high-cost patients, where high-risk includes those patients susceptible to failure with respect to PROs in the setting of appropriate alignment targets.[8] In some series of ASD surgeries, only 50% of patients improved beyond the MCID threshold, indicating the need for predictive modeling to identify those likely to benefit from surgery.[9,10] Long-standing multicenter registries of ASD patients offer a unique opportunity for the application of predictive analytics to model PROs after surgery. The purpose of this study was to develop predictive models of patient-reported clinical improvement and the likelihood of that improvement meeting or exceeding the MCID for the commonly utilized PROs in ASD surgery.

MATERIALS AND METHODS

Patient Population

Two independent, prospective multicenter ASD registries were merged. Inclusion criteria were: age ≥18 years, plan for surgical treatment, and available 2-year follow-up. ASD was defined by the presence of at least one of the following: scoliosis ≥20°, sagittal vertical axis (SVA) ≥5 cm, pelvic tilt ≥25°, or thoracic kyphosis ≥60°. Institutional Review Board approval was obtained at all sites (11 in the United States, 2 in Spain, 2 in Turkey, 1 in France, and 1 in Switzerland). Enrollment began the 16th of September 2008 in the United States and the 27th of July 2009 in Europe.

Modeling Predictors

Standard demographic variables included patient age, sex, height, weight, and number of previous spine surgeries. Full-length free-standing antero-posterior and lateral spine radiographs were analyzed using validated software (Spineview, ENSAM Laboratory of Biomechanics, Paris)[11] for assessment of radiographic parameters. Radiographic measures were performed at central locations using standard techniques.[12] Radiographic measurements included standard measures of ASD including C7 SVA, pelvic incidence, and pelvic incidence lumbar lordosis mismatch.

PROs included the ODI,[13] SRS22r,[14] and the Optum SF-36v2 Health Survey.[15] The ODI consists of 10 questions, scored from 0 to 100, where 100 indicates the worst disability possible. For the purpose of our study, ODI was inverted so that higher scores indicate better health to gain comparability with other instruments. The SRS-22r is a disease-specific (spine deformity) instrument, scored on a scale from 1 to 5, where 5 indicates perfect health. It is comprised of a summary score and four health domains: pain, self-image, function, and mental health, as well as a satisfaction score. The SF-36vs provides Physical Component Summary (PCS) and Mental Component Summary (MCS) scores, where a higher score indicates better health. PRO data were collected at baseline, 1 and 2 years postoperatively.

Surgical data collected included surgical approach, number of vertebral levels from the upper-most instrumented vertebrae to the lower-most instrumented vertebrae, use of pelvic fixation, operative time and estimated blood loss. In addition, use of interbody fusion was collected and whether it was performed as a transforaminal lumbar interbody fusion or anterior lumbar interbody fusion. Osteotomies were collected, including Smith-Peterson osteotomies, pedicle subtraction osteotomies, and vertebral column resections.

Minimal Clinically Important Difference

MCID values for the SRS-22r instrument have been proposed by Crawford *et al*[16] from a separate and distinct registry: Activity 0.6, Pain 0.4, Self-Image 1.2, Subscore 0.4. ODI, and SF-36 PCS MCID values were obtained from lumbar spine disease studies: ODI 13, SF-36 PCS 4.9.[8] PROs are, however, subject to measurement error.[17–20] Thus, we estimated a 10% measurement error to account for inaccuracy in MCID likelihood estimates.

Statistical Analysis

In total, 75 variables were used in the training of the regression models. Eight PRO domains (ODI, SRS-22r Function/Mental Health/Pain/Self-Image/Subscore, SF-36 PCS/MCS) served as the outcomes of interest at 1 and 2 years postoperatively. For baseline time horizons, both preoperative and immediate postoperative models were fitted. Time horizons of baseline to 1 year and baseline to 2 years trained eight separate PRO prediction models, providing 32 predictions per patient. Figure 1 presents the basic flowchart of the time horizons. Eight modeling algorithms were used: ordinary least squares,[21] ordinary least squares with partitions,[22] elastic net,[23] gradient boosting machines,[24] extreme gradient boosting tree,[25] extreme gradient boosting linear,[26] random forest,[27] and generalized linear modeling.[28] Thus, 256 models were created for each patient, predicting the change in PROs over four-time horizons. Goodness of fit assessed accuracy for each model by the root mean squared error and mean average error (MAE). Model selection was based on the minimization of MAE for each time-horizon, with 32 models selected (a model for each time horizon and each PRO). Standard errors were computed with in-sample variance. External validation for training and testing was performed with 80%/20% test sets. The predicted change in PROs from baseline was compared against MCID for each PRO outcome. Values meeting or exceeding (change in PRO—MCID $\geq$ 0) were considered surgical success. The confidence intervals of each prediction were used to assess the likelihood of both an improvement over the null and an improvement over the defined MCID values. Supplementary Table 1, http://links.lww.com/BRS/B421 presents a detailed TRIPOD[29] assessment of the modelling.

RESULTS

Of 1612 patients, 570 (35.5%) met inclusion criteria, with 743 (46.1%) excluded since they were not yet eligible for 2-year follow-up due to on-going database enrollment. Table 1 presents the main descriptive statistics for assessed patient parameters, including demographics, radiographic measures, and PROs. The mean patient age was 56.8 years, and women represented 78.8% of the patients. Kernel density plots for PROs at baseline, 1 year, and 2 years are shown in Figure 2. The rightward shift from baseline to 1 and 2 years indicates improvement in PROs after surgery. The general overlap of 1-year and 2-year values indicates reasonably static PROs between these time points (Table 2).

Final models for each time horizon and PRO were chosen based on goodness of fit, as assessed by the MAE. Goodness of fit/accuracy was considered excellent with MAE less than or equal to MCID for all PRO models. Measures of precision for the models are found in Figures 3 and 4. Models for ODI change were most precise, while SRS-22r pain was least precise between baseline and 1-year and 2-year follow-ups. For the 32-models, the average variable importance is shown in Table 3. Baseline PROs were most important for predicting final PRO values. Age at baseline was the most important objective, patient-level variable. Hypertension, arthritis, and depression were the comorbid conditions of primary importance in the models. Major coronal Cobb measurement was among the least important variable in overall model prediction.

Considering a 10% measurement error, a maximum R^2 value of 0.81 is expected. Our estimates reflect a range of

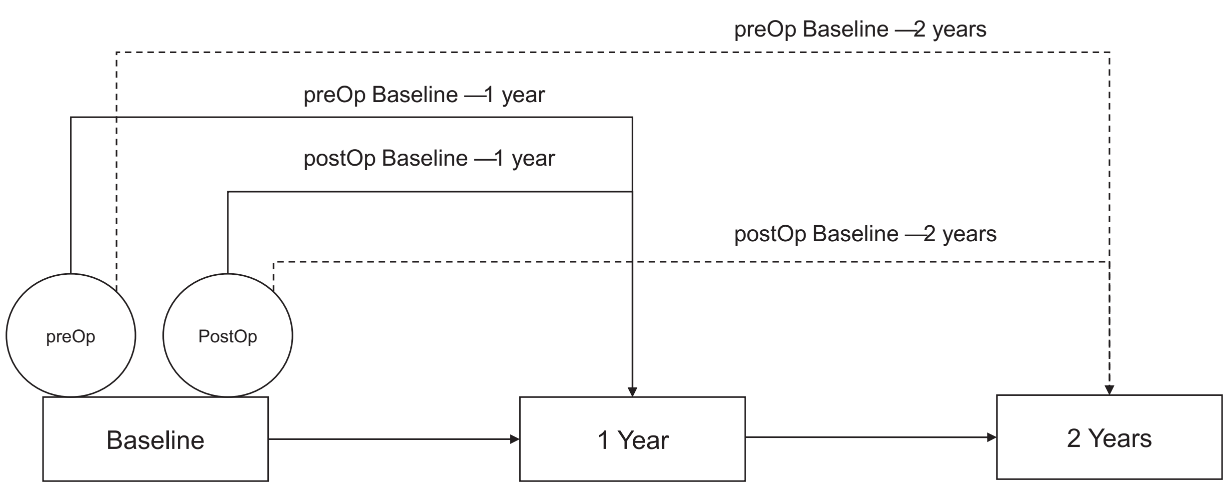

Figure 1. Predictive modeling flowchart.

TABLE 1. Patient Parameters: Demographics, Radiographic Measures, and Patient-reported Outcomes Scores for 570 Adult Spinal Deformity Patients

Variable	Mean or %	SD	Median	Min	Max	SE
Preoperative baseline						
Sex (% women)	79					1.7
Height (cm)	162.55	9.61	162	104.5	195.6	0.4
Weight (kg)	70.31	16.03	68	40	158.8	0.67
Age (yr)	56.75	16.32	59.93	18.05	86.42	0.68
Sagittal alignment (SVA, cm)	55.64	69.55	44.89	−83.73	314.09	2.91
Coronal alignment (GCA, cm)	31.24	31.4	23.86	0	288.95	1.32
Major curve Cobb angle (°)	40.27	21	38.4	0	123.7	0.88
Pelvic tilt (°)	23.28	11.03	23.39	−12.66	67.9	0.46
Inverse ODI score*	57.55	18.89	56	8	100	0.79
SRS-22r function score	2.98	0.88	3	1	5	0.04
SRS-22r mental health score	3.31	0.93	3.25	1	5	0.04
SRS-22r pain score	2.49	0.88	2.4	1	5	0.04
SRS-22r self-image score	2.43	0.73	2.4	1	4.6	0.03
SRS-22r subtotal score	2.8	0.66	2.82	1.09	4.5	0.03
SF36v2 MCS score	44.53	13.19	45.62	10.59	73.53	0.55
SF36v2 PCS score	33.38	9.85	32.29	8.65	63.58	0.41
1-yr postoperative						
Sagittal alignment (SVA, cm)	23.83	51.38	18.99	−114.08	181.03	2.15
Coronal alignment (GCA, cm)	4.98	30.23	8.66	−162.17	113.25	1.27
Major curve Cobb angle (°)	9.09	23.43	10.96	−77.5	100.2	0.98
Pelvic tilt (°)	20.83	9.6	20.79	−15.29	61.13	0.4
Inverse ODI score*	72.44	18.94	74.78	16	100	0.79
SRS-22r function score	3.5	0.9	3.6	1	5	0.04
SRS-22r mental health score	3.75	0.88	4	1	5	0.04
SRS-22r pain score	3.45	0.96	3.6	1	5	0.04
SRS-22r self-image score	3.64	0.86	3.6	1	5	0.04
SRS-22r subtotal score	3.63	0.74	3.73	1.14	5	0.03
SF36v2 MCS score	49.34	12.32	52.58	10.48	73.95	0.52
SF36v2 PCS score	40.98	10.03	41.42	11.47	66.45	0.42
2-yr postoperative						
Sagittal alignment (SVA, mm)	30.36	54.23	24.54	−109.2	237.1	2.34
Coronal alignment (GCA, mm)	23.37	19.57	18.18	0	114.8	1.34
Major curve Cobb angle (°)	20.59	16.35	17	0	94.5	0.87
Pelvic tilt (°)	20.87	10.11	21	−6.3	61.87	0.45
Inverse ODI score*	72.51	19.98	76	17.78	100	0.84
SRS-22r function score	3.53	0.95	3.6	1.2	5	0.04
SRS-22r mental health score	3.72	0.92	3.8	1	5	0.04
SRS-22r pain score	3.44	1.06	3.6	1	5	0.04
SRS-22r self-image score	3.57	0.91	3.6	1	5	0.04
SRS-22r subtotal score	3.6	0.81	3.68	1.23	5	0.03
SF36v2 MCS score	48.57	12.43	51.33	7.32	70.87	0.52
SF36v2 PCS score	40.95	11.06	41.12	13.01	71.61	0.46

*For the purposes of computational analyses, ODI scores are presented as inverse values, with scores of 0 and 100 reflecting the most and least degrees of disability, respectively.

GCA indicates global coronal alignment; MCS, mental component score; ODI, Oswestry disability index; PCS, physical component score; SD, standard deviation; SE, standard error, SVA, C7-S1 sagittal vertical axis; SRS, Scoliosis Research Society.

24.7% to 69.1% in relative precision depending upon time horizon and outcome of interest. S2 supplementary material Table 1, http://links.lww.com/BRS/B421 presents additional descriptive statistics and robustness checks.

Simulation Study

Table 4 presents the probabilities for overall improvement in PROs and improvement that meets or exceeds PRO MCID for a hypothetical patient: a 60-year-old woman with one prior spine surgery, no comorbidities,

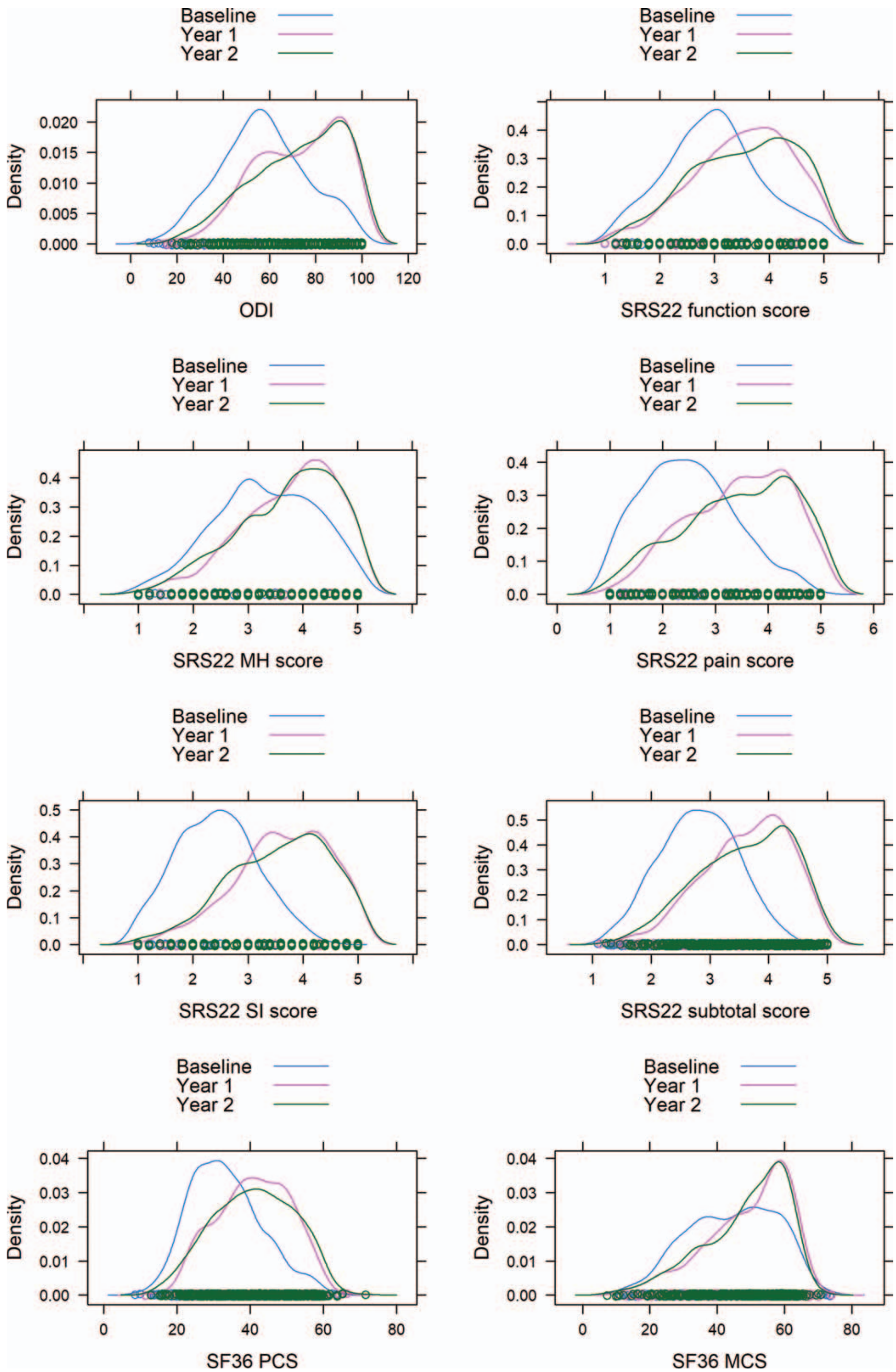

Figure 2. Kernel density plots for patient-reported outcomes scores at baseline and at 1-year and 2-year follow-ups.

employed, 168 cm tall, 64 kg, C7SVA +72 mm, 12° major Cobb, 31° pelvic tilt. The hypothetical surgery is a 10-level, all-posterior surgery with a single pedicle subtraction osteotomies, iliac instrumentation, and no interbody fusions.

DISCUSSION

Predictive analytics are employed in many fields where large datasets are available. Maturation of two large, observational cohorts now makes these types of analyses possible for ASD. In a value-driven healthcare system, predictive

TABLE 2. Surgical Parameters for 570 Adult Spinal Deformity Patients

Variable	Mean or %	SD	Median	Min	Max	SE
Previous spine surgery (%)	41.6					2
Any major complication (%)	36.2					2
Total surgical time (min)	345.74	134.47	342	43	802	5.63
Estimated blood loss (cc)	1660.16	1465.51	1300	20	12,200	61.38
Number of fused vertebral levels	10.7	4.21	10	0	23	0.18
Use of pelvic fixation (%)	59.5					2
Use of interbody fusion (%)	54.4					2
Use of Smith-Petersen osteotomy (%)	50.0					2
Use of 3-column osteotomy (%)	20.8					1.7
Length of hospitalization (days)	8.88	6.49	7	0	83	0.27

SD indicates standard deviation; SE, standard error.

analytics will be invaluable as they may assist with identification of patients at high risk for complications and surgical failures. As value is often defined as changes in PROs divided by the cost of care, it will be necessary to identify patients optimized for success from a PRO perspective. Similarly, it will be necessary to identify those patients at risk for PRO failure and to at least optimize these patients by identifying the patient-level variables that may be modified to improve outcomes. Finally, surgical plans developed using predictive models in concert with baseline patient data will offer the opportunity for customized care in ASD.

The purpose of this study was to use two large, observational cohort studies of ASD patients treated with surgery to develop prediction models for surgical success, where success was defined as a PRO change meeting or exceeding the MCID for the SF-36v2, ODI, or SRS-22r. The prediction models were created using eight modeling methods to identify the most accurate estimate for each outcome of interest with four distinct time horizons. The composite of these models created a prediction tool for each patient. We found that baseline PROs are the most important component of predicting change in PROs and, thus, success from this perspective. Age is the most important nonmodifiable patient-level characteristic in the prediction tool. There are important points illustrated in the simulation case, where a delay to surgery accompanied by a decline in PRO does not necessarily negatively affect the likelihood of achieving a clinically relevant PRO improvement. Future work should include efforts to define the floor and ceiling effects of surgery, as there may be an optimal point of intervention to maximize outcomes and minimize any detrimental effects of surgery, such as decline in function due to a multilevel fusion.

Despite the frequent attention paid to sagittal plane alignment and correction, postoperative alignment was not an important component of the final model. This warrants further investigation, given our emerging understanding of age-adjusted alignment goals, lumbopelvic harmony, and the subjective nature of PRO.[30–32] Not surprisingly, comorbid conditions associated with other health-related quality of life complaints such as hypertension and arthritis were important in the model, though less so than baseline PROs. This has multiple implications for PRO selection and model selection. For example, a disease-specific instrument, such as the SRS-22r, may not capture disability due to hip arthritis, while a general measure of health, such as the SF-36v2, may be sensitive to the detrimental change associated with osteoarthritis of the hip. An aging patient with hip osteoarthritis may be predicted to do poorly if the primary outcome is SF-36 PCS, while the SRS-22r may predict a successful outcome due to different PRO perspective. This emphasizes the fact that patient expectations should be discussed and defined in customized care.

Numerous prior studies in ASD surgery have investigated factors associated with good outcomes.[33–38] Factors associated with worse outcomes are often termed "predictors" or "risk factors" when they are merely associations. These associations are determined from multivariable regression analyses, which allow for control of other variables within the regression models. Interpretation of these results may be difficult as they are usually presented as odds ratios or relative risks, which require comprehension of the risk within the reference group.[39] Furthermore, interpretation of any single variable is not a situation surgeons face in the multifactorial decision-making process. That is, the risk estimated is conditional upon other variables included in the model and their values. Finally, the number of data points available limits the number of covariates within models. In many cases, models are limited to a finite number of covariates due to small enrollment numbers and, thus, few covariates are included and controlled, limiting clinical applicability. As seen here, the relative lack of importance of frequently emphasized patient demographic data, such as depression, weight, and amount of sagittal plane malalignment, strengthens the argument for the use of complex multivariable prediction models.

Prediction algorithms have been explored in other fields of medicine, with outcome variables ranging from supply chain needs to complications related to care.[40–43] Limitations of these algorithms come from the data that creates them and subsequent iterations of these models, including

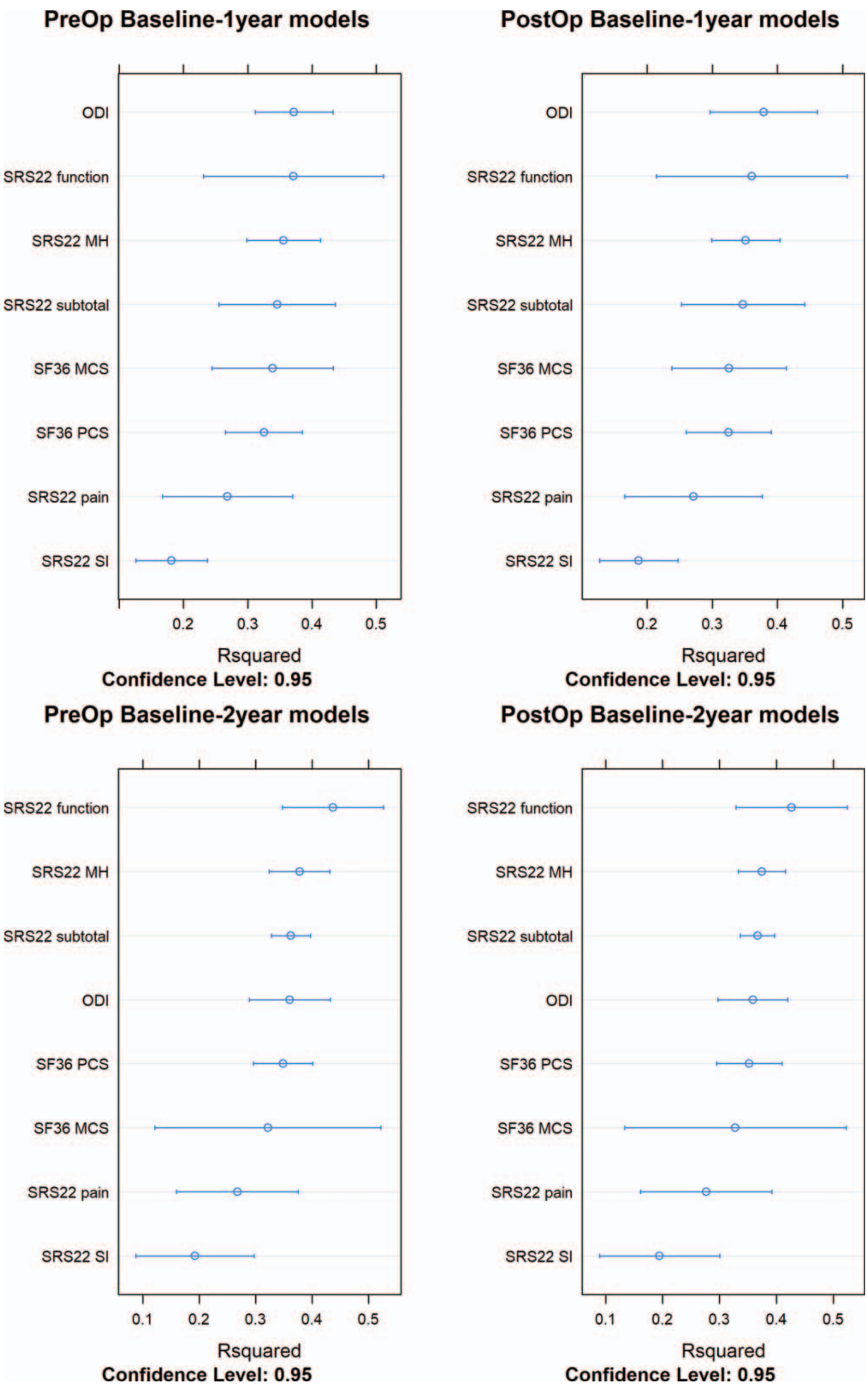

Figure 3. Test set precision of patient-reported outcomes measured with estimates at four-time horizons, including with the baseline set as either the preoperative time point or immediately following surgery.

ours, will follow as more data are added to the system. The limitations are most profound when estimating risk for complications or some deleterious outcome, due to individual characteristics of any single patient, surgeon, and surgery. The addition of models, such as these, in an effort to identify those likely to benefit from an intervention may be a critical step in providing evidence-based care in the future.[44]

The number of patients enrolled in these two registries may appear low, though one must consider both the quantity and quality of data provided. The 570 patients studied contributed many thousands of data points, all collected prospectively into databases designed for clinical research. "Big data" in medicine often means analysis of administrative databases, not designed for clinical research and missing

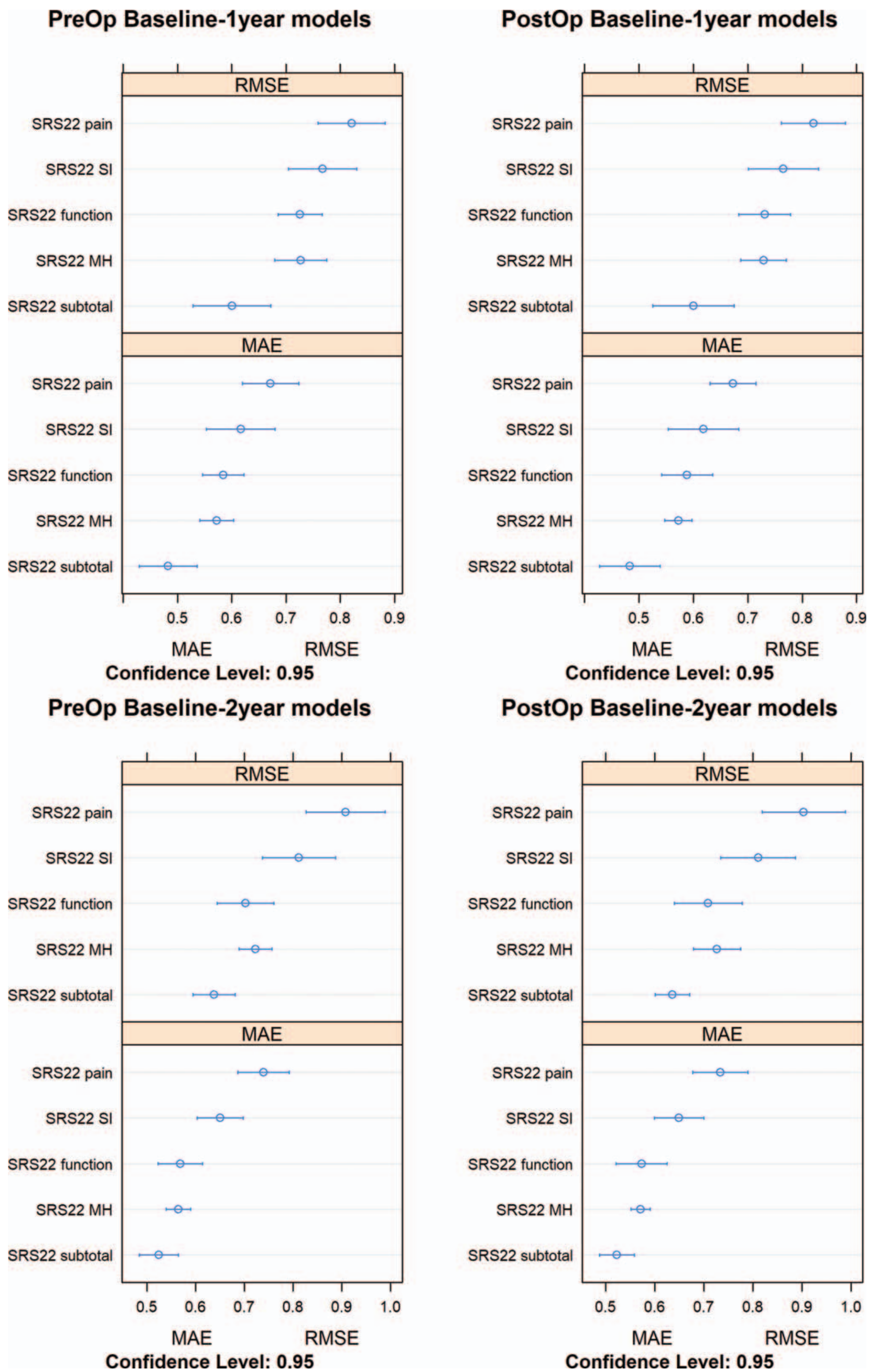

Figure 4. Test set precision of the SRS-22r scores with estimates at four time horizons, including with the baseline set as either the preoperative time point or immediately following surgery.

essential data to improve the external validity of the findings. We feel confident that our combined dataset offers a broad exposure to ASD patients seeking surgical treatment and offers good external validity. This is supported by the low importance of the specific of surgeon in our models. Modeling of subjective data, such as PROs, can be difficult due to unmeasurable data which may affect patient reporting. However, as PROs are a necessary feature of value-driven healthcare, we feel it is important to include them for these analyses.

TABLE 3. Average Relative Variable Importance for Estimation of Postoperative Patient-reported Outcomes

Names	Importance
SRS22 subtotal score baseline	12.57%
ODI score baseline	10.32%
SF36v2 MCS score baseline	9.26%
SRS22 MH score baseline	8.31%
SF36v2 PCS score baseline	7.74%
SRS22 function score baseline	7.33%
SRS22 pain score baseline	5.73%
SRS22 SI score baseline	3.82%
Age baseline	4.44%
Level of physical labor null	3.23%
Prior spine surgery yes	3.11%
Pelvic tilt baseline	2.45%
Hypertension baseline	2.35%
Arthritis baseline yes	2.31%
Fused vertebrae	2.26%
Height cm baseline	2.26%
Sagittal balance baseline	2.25%
Number of levels between UIV and LIV	2.22%
Major curve Cobb angle baseline	2.12%
Depression baseline yes	2.00%
Weight kg baseline	1.97%
SPOs	1.95%

Only the average top 22 predictors are included in the table. Relative importance measure: mean decrease in node impurity.

LIV indicates lower-most instrumented vertebra; MCS, mental component summary; ODI, Oswestry disability index; PCS, physical component summary; SPOs, Smith-Petersen Osteotomies; SRS, Scoliosis Research Society; UIV, upper-most instrumented vertebra.

CONCLUSION

Customized ASD care is a concept in its infancy. Substantial efforts have worked toward defining appropriate alignment in ASD and have led to the development of custom implants and surgical plans. To refine customized care in ASD, prediction models are needed to improve the shared decision-making process and to aid surgeons in surgical planning. Predictive models to estimate change in PROs, likelihood of complications, and costs of care are all needed to maximize the likelihood of success in a value-driven healthcare system. We have shown here that models can be created to accurately predict the likelihood of achieving a change equal to or greater than the MCID for three commonly used PRO instruments in ASD. This is the first step in a long-range goal of providing patients and surgeons with tools to be used real time in the clinic to improve individual and overall clinical outcomes in ASD surgery.

➢ Key Points

- ❑ A predictive model on the likelihood of overall and over MCID improvement following ASD surgery was developed and tested with the most used HRQoL instruments.
- ❑ A total of 570 patients with 2-year follow-up were included, 80% of the sample was used for training and the other 20% was used for testing.
- ❑ Goodness of fit measured in the test sample was between 20% and 45%, indicating a successful fit.

TABLE 4. Probabilities of Overall Improvement and Improvement Meeting or Exceeding MCID for ODI, SRS-22r, and SF-36 PCS/MCS

	HRQL Instrument	Baseline Score	Baseline Probabilities	Depression and Hypertension	10% Improve in Physical HRQL Scores	Waiting 5 yr With a Reduction of 10% HRQL	Waiting 10 yr With a Reduction of 20% HRQL	Postop SVA 0 mm	Postop SVA 50 mm	Postop SVA 100mm	Number of Levels From 10 to 15	Range of Variation Across Options
(1) Overall improvement	ODI	68	54.4%	55.1%	43.5%	67.6%	72.6%	54.4%	54.4%	54.4%	54.4%	29.1%
(2) Improvement over MCID	ODI	68	38.4%	39.1%	28.4%	52.0%	57.7%	38.4%	38.4%	38.4%	38.4%	29.3%
(1) Overall improvement	SRS22 function	3.2	65.0%	66.1%	51.4%	76.4%	83.5%	65.0%	65.0%	65.0%	65.0%	32.1%
(2) Improvement over MCID	SRS function	3.2	35.4%	36.5%	23.4%	48.3%	58.4%	35.4%	35.4%	35.4%	35.4%	35.0%
(1) Overall improvement	SRS22 MH	2.4	80.6%	80.6%	82.0%	86.0%	90.3%	80.6%	80.6%	80.6%	78.6%	11.7%
(2) Improvement over MCID	SRS22 MH	2.4	64.7%	64.7%	66.6%	72.4%	79.2%	64.7%	64.7%	64.7%	62.0%	17.2%
(1) Overall improvement	SRS22 pain	2	61.4%	63.7%	67.1%	76.2%	82.6%	61.4%	61.4%	61.4%	62.2%	21.2%
(2) Improvement over MCID	SRS22 pain	2	46.2%	48.6%	52.3%	62.9%	71.0%	46.2%	46.2%	46.2%	47.0%	24.8%
(1) Overall improvement	SRS22 SI	2.4	81.5%	81.5%	83.7%	84.3%	86.9%	81.5%	81.5%	81.5%	81.5%	5.4%
(2) Improvement over MCID	SRS22 SI	2.4	34.6%	34.6%	37.8%	38.8%	43.3%	34.6%	34.6%	34.6%	34.6%	8.7%
(1) Overall improvement	SRS22 subtotal	2.38	69.8%	72.1%	74.7%	84.8%	91.7%	69.8%	69.8%	69.8%	70.7%	21.9%
(2) Improvement over MCID	SRS22 subtotal	2.38	45.8%	48.4%	51.7%	65.6%	77.7%	45.8%	45.8%	45.8%	46.8%	31.9%
(1) Overall improvement	SF36v2 MCS	22.18	95.6%	95.4%	95.6%	97.0%	98.0%	95.6%	95.6%	95.6%	95.6%	2.6%
(2) Improvement over MCID	SF36v2 MCS	22.18	81.5%	81.0%	81.5%	85.9%	89.5%	81.5%	81.5%	81.5%	81.5%	8.5%
(1) Overall improvement	SF36v2 PCS	39.66	44.3%	45.3%	34.6%	60.4%	63.4%	44.3%	44.3%	44.3%	43.5%	28.8%
(2) Improvement over MCID	SF36v2 PCS	39.66	17.4%	18.0%	11.6%	29.7%	32.4%	17.4%	17.4%	17.4%	16.9%	20.8%

Patient: Female with one prior spine surgery, with loss of balance, without comorbidities, employed, steady gait, 60 years old, 167.6 cm height, 63.9 kg weight, 71.68 sagittal balance, 11.8° of major curve cobb angle, 30.59° pelvic tilt. Surgery: pelvic fixation, 7 fused vertebrae, posterior instrumentation, 1 PSOs, 0 SPOs, no PLIF, no TLIF, no ALIF, 10 levels between UIV and LIV and no interbody fusion.

- Patients with worse preoperative baseline PROs were the most likely to achieve clinically relevant improvements.

Supplemental digital content is available for this article. Direct URL citations appearing in the printed text are provided in the HTML and PDF version of this article on the journal's Web site (www.spinejournal.com).

References

1. Porter ME, Lee TH. From volume to value in health care: the work begins. *JAMA* 2016;316:1047–8.
2. Porter ME. What is value in health care?. *N Engl J Med* 2010;363:2477–81.
3. Glassman SD, Bridwell K, Dimar JR, et al. The impact of positive sagittal balance in adult spinal deformity. *Spine (Phila Pa 1976)* 2005;30:2024–9.
4. Wilke A, Mata R. Cognitive Bias. In: *Encyclopedia of Human Behavior*. Elsevier: 531–5.
5. Gehlbach H, Barge S. Anchoring and adjusting in questionnaire responses. *Basic Appl Soc Psych* 2012;34:417–33.
6. Gum JL, Bridwell KH, Lenke LG, et al. SRS22R appearance domain correlates most with patient satisfaction after adult deformity surgery to the sacrum at 5-year follow-up. *Spine (Phila Pa 1976)* 2015;40:1297–302.
7. Gum JL, Glassman SD, Carreon LY. Clinically important deterioration in patients undergoing lumbar spine surgery: a choice of evaluation methods using the Oswestry Disability Index, 36-Item Short Form Health Survey, and pain scales. *J Neurosurg Spine* 2013;19:564–8.
8. Copay A, Subach B, Glassman S, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541–6.
9. Nunez-Pereira S, Vila-Casademunt A, Domingo-Sabat M, et al. Impact of early unanticipated revision surgery on health-related quality of life after adult spinal deformity surgery. *Spine J* 2017;18:926–934.
10. Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff* 2014;33:1123–31.
11. Lafage R, Ferrero E, Henry JK, et al. Validation of a new computer-assisted tool to measure spino-pelvic parameters. *Spine J* 2015;15:2493–502.
12. O'Brien MF, Group SDS. Radiographic Measurement Manual. Medtronic Sofamor Danek USA. Available at: https://books.google.es/books?id=If_0swEACAAJ.2004. Accessed January 9, 2018.
13. Fairbank JC, Pynsent PB. The Oswestry disability index. *Spine (Phila Pa 1976)* 2000;25:2940–52.
14. Asher MA, Lai SM, Glattes RC, et al. Refinement of the SRS-22 health-related quality of life questionnaire function domain. *Spine (Phila Pa 1976)* 2006;31:593–7.
15. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;306:1437–40.
16. Crawford CH, Glassman SD, Bridwell KH, et al. The minimum clinically important difference in SRS-22R total score, appearance, activity and pain domains after surgical treatment of adult spinal deformity. *Spine (Phila Pa 1976)* 2015;40:377–81.
17. Butler JS, Burkhauser RV, Mitchell JM, et al. Measurement error in self-reported health variables. *Rev Econ Stat* 1987;69:644–50.
18. Takada K, Takahashi K, Hirao K. Measurement error in the Liebowitz Social Anxiety Scale: results from a general adult population in Japan. *Int J Psychiatry Clin Pract* 2018;22:289–95.
19. An Y, Hu Y. Well-posedness of measurement error models for self-reported data. *J Econom* 2012;168:259–69.
20. Lim S, Wyker B, Bartley K, et al. Measurement error of self-reported physical activity levels in New York City: assessment and correction. *Am J Epidemiol* 2015;181:648–55.
21. Schumacker R, Monahan M, Mount R. A comparison of OLS and robust regression using S-PLUS. *Mult Linear Regres* 2002;28:10–3.
22. Park Y, Kim D, Kim S. Robust regression using data partitioning and M-estimation. *Commun Stat Simul Comput* 2012;41 [Epub ahead of print].
23. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Statist Method)* 2005;67:301–20.
24. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–232.
25. Chen T, He T. Xgboost: extreme gradient boosting. *R Packag version 04-2*.
26. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM:785–94.
27. Breiman L. University of C. Random forest. *Mach Learn* 1999;45: 1–35.
28. Song L, Langfelder P, Horvath S. Random generalized linear model: A highly accurate and interpretable ensemble predictor. *BMC Bioinformatics* 2013;14:5.
29. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or disagnosis (TRIPOD): explanantion and elaboration. *Ann Intern Med* 2015;162:W1–73.
30. Lafage R, Schwab F, Challier V, et al. Defining spino-pelvic alignment thresholds should operative goals in adult spinal deformity surgery account for age?. *Spine (Phila Pa 1976)* 2016;41:62–8.
31. Meyer KB, Clayton KA. Measurement and analysis of patient-reported outcomes. *Methods Mol Biol* 2009;473:155–69.
32. Yilgor C, Sogunmez N, Boissiere L, et al. Global alignment and proportion (GAP) score. *J Bone Joint Surg* 2017;99:1661–72.
33. Carreon LY, Glassman SD, Shaffrey CI, et al. Predictors of health-related quality-of-life after complex adult spinal deformity surgery: a scoli-RISK-1 secondary analysis. *Spine Deform* 2017;5:139–44.
34. Gum JL, Carreon LY, Stimac JD, et al. Predictors of oswestry disability index worsening after lumbar fusion. *Orthopedics* 2013;36:e478–83.
35. Lin T, Meng YC, Li TB, et al. Predictors of postoperative recovery based on health-related quality of life in patients after degenerative lumbar scoliosis surgery. *World Neurosurg* 2018;109:e539–45.
36. Mmopelwa T, Ayhan S, Yuksel S, et al. Analysis of factors affecting baseline SF-36 Mental Component Summary in Adult Spinal Deformity and its impact on surgical outcomes. *Acta Orthop Traumatol Turc* 2018;52:179–84.
37. Raad M, Neuman BJ, Jain A, et al. The use of patient-reported preoperative activity levels as a stratification tool for short-term and long-term outcomes in patients with adult spinal deformity. *J Neurosurg Spine* 2018;29:68–74.
38. Sharma A, Tanenbaum JE, Hogue O, et al. Predicting clinical outcomes following surgical correction of adult spinal deformity. *Neurosurgery* 2019; 84:733–740.
39. Cohen L. An introduction to the philosophy of induction and probability. *Mind* 1989;99:313–5.
40. Dinov ID, Heavner B, Tang M, et al. Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS One* 2016;11:0157077.
41. Cichosz SL, Johansen MD, Hejlesen O. Toward big data analytics: review of predictive models in management of diabetes and its complications. *J Diabetes Sci Technol* 2015;10:1–8.
42. Basu RK, Gist K, Wheeler DS. Improving acute kidney injury diagnostics using predictive analytics. *Curr Opin Crit Care* 2015;21:473–8.
43. Stefanovic N. Proactive supply chain performance management with predictive analytics. *ScientificWorldJournal* 2014;2014: 528917.
44. Sniderman AD, D'Agostino RB, Pencina MJ. The role of physicians in the era of predictive analytics. *JAMA* 2015;314:25–6.